Regression Algo:

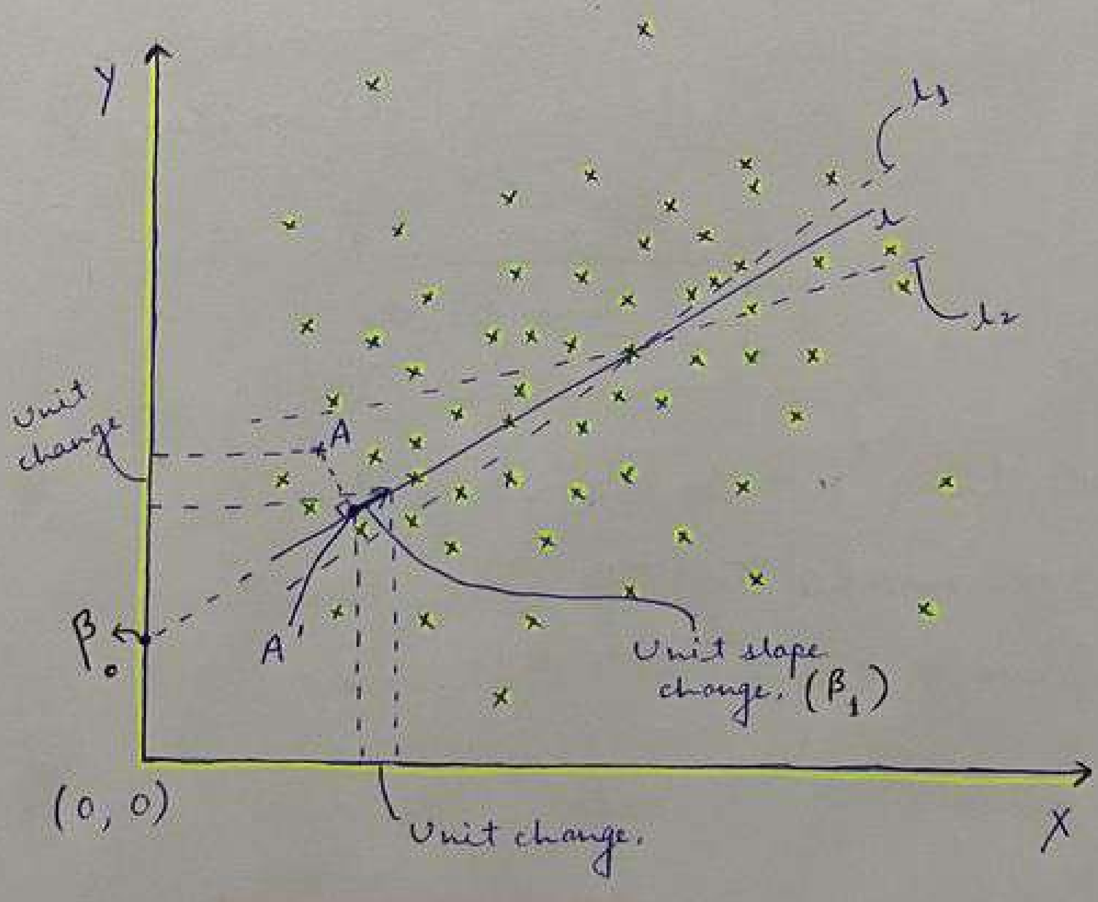
Simple linear Regression:

It's used to estimate the relationship b/w two quantitative variables.

we find the best fit line, such that the distance of error b/w datapoints and the line is minimal.

Predicted points lies on the line which is being trained by the model.

[× Data points.].



Now,

$\boxed{\text{Error or residual = Distance of actual point - Predicted point dim. on the best fit line.}}$

From previous graph we have actual point A and when A predicted on the line centre is A'.

Thus,

$$\text{Error} = \dim(A) - \dim(A')$$

$$= A(x_i, y_i) - A'(x_i, y_i)$$

Equation of st. line:

$$y = mx + c$$

$$\Rightarrow \boxed{y(x) = \beta_1 x + \beta_0}$$

β_1 : Slope [for any predicted point, unit movement cause of Best fit line bcz of Unit movements in X and y axis.]

β_0 : Intercept [value of y when $x=0$]

y : Predicted value w.r.t independent Variable x.

Methods to find best fit line:

Mean squared Error: (M.S.E)

We measure the average of square of error of the difference b/w estimated value and the actual value.

M.S.E:

$$\boxed{J(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Above formulae is also known as cost function.

- n : No.s of datapoints.
- y_i : Observed value of variable
- $\hat{y}_i$: Predicted value of variable

To get best fit line we need to minimize the error / residual.

Thus we need to minimize the cost function.

Consider below example and assume best fit line passes through origin.

$\Rightarrow \beta_0 = 0$

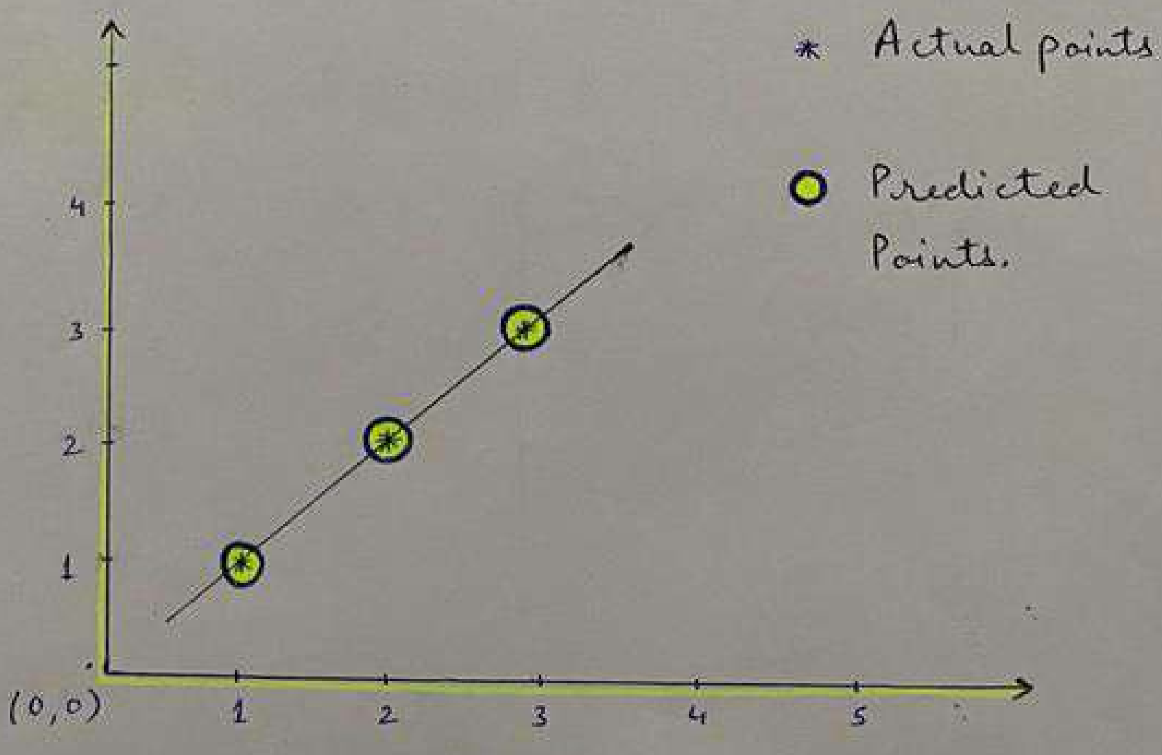
Thus, we have, and also consider, $\beta_1 = 1$

$$\beta_1(x) = y(x) = x$$

x	y	$\beta_1(x) = y(x)$
1	1	1
2	2	2
3	3	3

Training data points   Predicted Points.

Plotting graph for both data.



Calculating $J(\beta_0, \beta_1)$

$$= \frac{1}{3}(0 + 0 + 0) \qquad [y_i - \hat{y}_i = 0]$$

Thus, Proved when the error / residual is minimum we have the best fit line.

Calculating, $J(\beta_0, \beta_1)$ for various value of β_1 using above data points.

$\beta_0 = 0.5$ $\quad [y(x) = \beta_1(x)]$

$y(1) = .5$

$y(x) = 1$

$y(3) = 1.5$

$$J(\beta_1) = \frac{1}{3}\left[(.5)^2 + 1^2 + (-1.5)^2\right]$$

$$= \frac{3.5}{3} = 1.16$$

- - - - - - - - - - - -

$\beta = 1.5$ $\quad [y(x) = \beta_1(x)]$

$y(1) = 1.5$

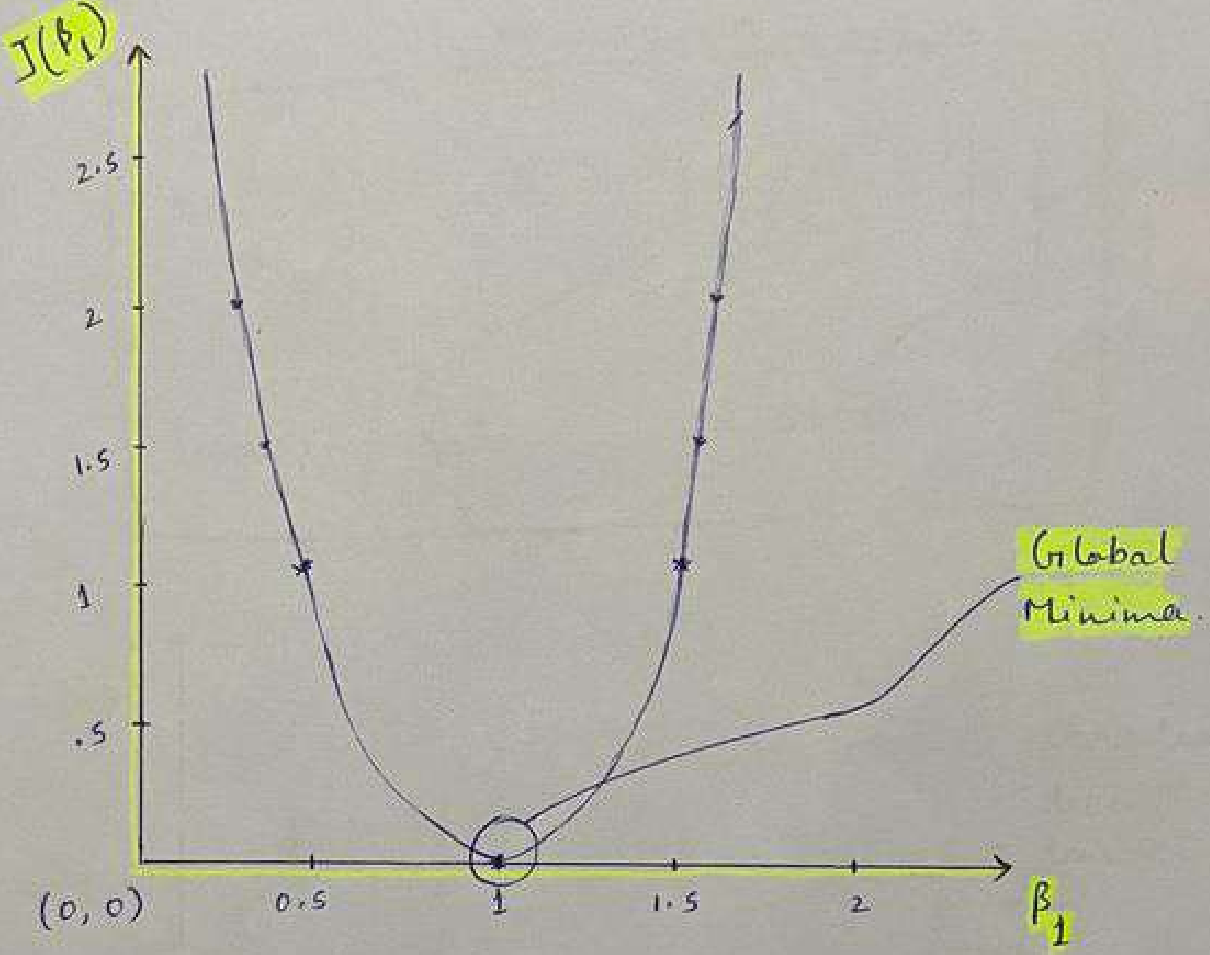
$y(2) = 3$

$y(3) = 4.5$

$$J(\beta_1) = \frac{1}{3}\left[(.5)^2 + 1^2 + (1.5)^2\right]$$

$$= 1.16$$

Plotting graph with value of β_1 and $J(\beta_1)$ considering various value of β_1. we get.



we get parabala when we draw the graph. Curve is known as Gradient decent Curve.

In practical it's not possible to keep changing value of β_1 to come to the global minima. So we use convergence Algorithm.

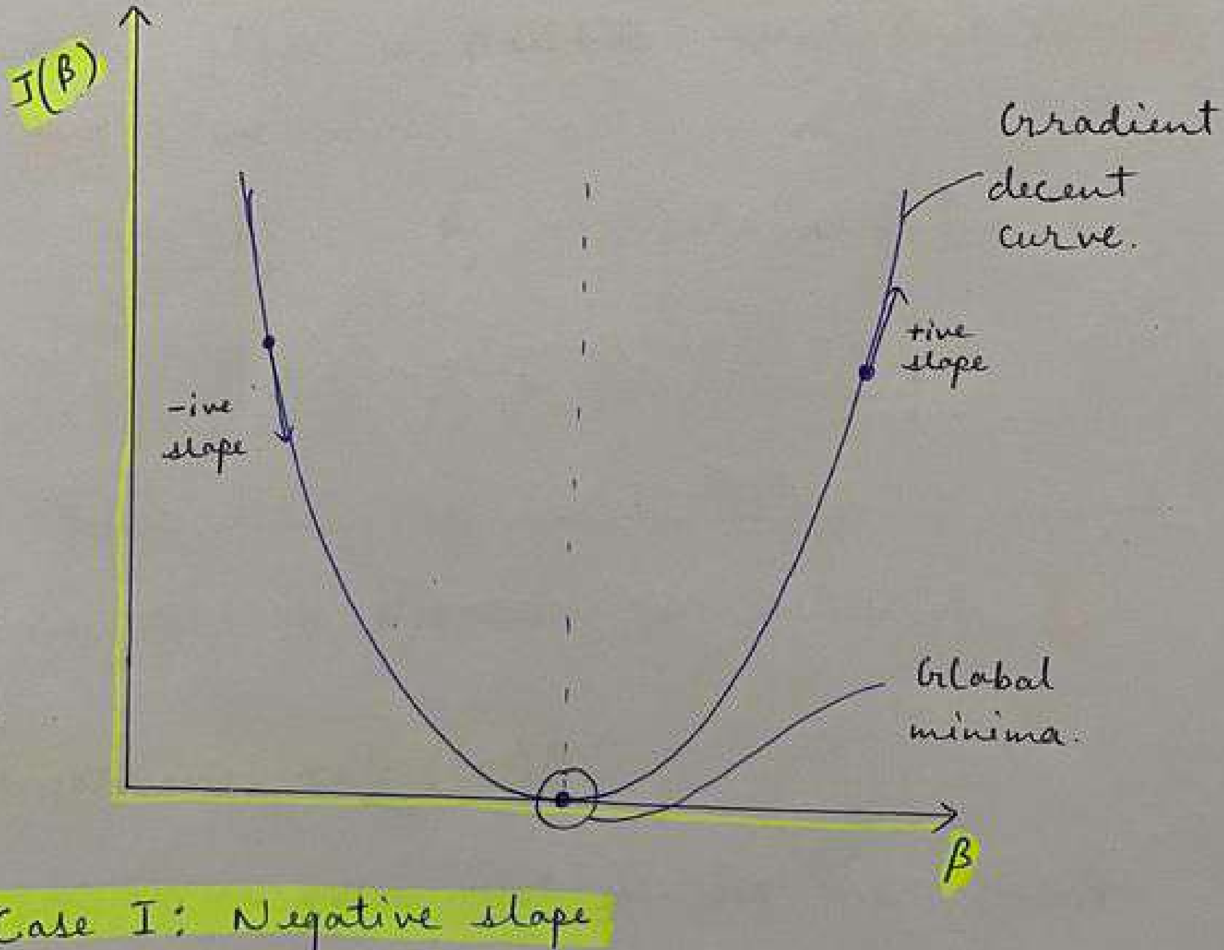
Convergence Algarithm.

We use to aptimige the change of β.

Repeat until convergence:

$$\left\{ \beta_j = \beta_j - \alpha \left[\frac{\partial J(\beta_j)}{\partial \beta_j} \right] \right\}$$

learning rate   Slape.



Case I: Negative slope

when we have -ive slape means we need to decrease the value of β in arder to reach to global minima. Slope is negative.

Thus, we have,

$$\beta_j = \beta_j - \alpha [\text{-ive slope}]$$

$$= \beta_j + [\text{Positive value}]$$

Mean while we need to increase the value of β to reach to global min.

Case II: Positive slope.

when slope is positive means β is increa-sing. we have,

$$\beta_j = \beta_j - \alpha [\text{+ive slope}]$$

$$= \beta_j - [\text{Positive value}]$$

Mean while to reach the global minima we need to decrease the value of β_j.

Learning rate (α):

It's parameter that decides the speed of convergence means how fast we can reach the global minima.